AD-A055 042 TEXAS UNIV AT AUSTIN CENTER FOR CYBERNETIC STUDIES F/G 12/1

DEGENERACY IN SPECIAL PURPOSE PRIMAL ALGORITHMS USED IN OBTAINI--ETC(U)

NOV 77 R D ARMSTRONG, A CHARNES, J P GODFREY N00014-75-C-0616

UNCLASSIFIED CCS-316 NL

1 OF 1

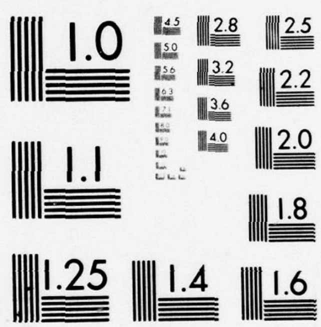
AD A055042

END

DATE FILMED

7-78

DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A055042

R FURTHER TRAN

12

BS

CENTER FOR CYBERNETIC STUDIES

The University of Texas
Austin, Texas 78712

78 06 06 055

(12)

Research Report CCS 316

DEGENERACY IN SPECIAL PURPOSE PRIMAL ALGORITHMS USED IN OBTAINING LEAST ABSOLUTE VALUE ESTIMATORS

by

R. D. Armstrong*
A. Charnes**
J. P. Godfrey***

November 1977

*Assistant Professor of Statistics, Department of General Business, The University of Texas, Austin, Texas 78712

**Director, Center for Cybernetic Studies, The University of Texas, Austin, Texas 78712

***Research Associate, Center for Cybernetic Studies, The University of Texas, Austin, Texas 78712

This research was partly supported by NSF Grant MCS77-00100 and by Project NR047-021, ONR Contracts N00014-75-C-0616 and N00014-75-C-0569 with the Center for Cybernetic Studies, The University of Texas. Reproduction in whole or in part is permitted for any purpose of the United States Government.

CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director
Business-Economics Building 203E
The University of Texas
Austin, Texas 78712
(512) 471-1821

Degeneracy in Special Purpose Primal Algorithms

Used in Obtaining Least Absolute Value Estimators

ABSTRACT

Efficient algorithms have been developed recently which utilize the specialized structure of the linear programming formulation for the problem of least absolute value estimation. These algorithms generally proceed in the direction of steepest descent along an edge of a convex polyhedral surface. However, we will show that the extreme point path of steepest descent may not be taken when degeneracy occurs. We will also present a criterion that determines the basic edge for steepest descent.

Keywords: L_1 Norm; Least Absolute Value; Linear Programming

1. Introduction

Least absolute value, or L_1 norm, estimates have long been considered as an acceptable alternative to least squares estimates. Fourier (See Darboux) appears to have been the first to examine the problem of obtaining such estimates. Later, Edgeworth (1887, 1888, 1923) also investigated the problem. The French mathematician de la Vallée Poussin (1911) proposed a solution procedure. In addition, algorithms were presented by Rhodes (1930) and by Singleton (1940). The labor involved in computing the least absolute value (LAV) estimators made the technique unpopular, and for a period of time most results were restricted to the cases of either one or two parameters. It was not until the work of Charnes, Cooper, and Ferguson (1955) that a practical procedure for obtaining LAV estimators was given. They demonstrated that the problem of minimizing the sum of absolute deviations could be reduced to an equivalent linear programming (LP) formulation.

With this development available to researchers, there has been a steadily increasing interest in LAV estimators and their properties in recent years. One of the main directions of this research was to find efficient, specialized algorithms which would solve the linear programming formulation. Of particular interest are the algorithms of Davies (1967) and Barrodale and Roberts (1973). These closely related algorithms use the special structure of the problem to perform "multiple pivots" before actually performing an iteration of the simplex algorithm. In this paper, we shall point out that these multiple pivoting techniques will not always give the extreme point path of steepest descent when degeneracy occurs. The possibility of cycling of bases is not a problem since it can be resolved by the perturbation technique of Charnes (1952).

2. Linear Programming Formulation

Mathematically, the least absolute value problem is that of finding estimates of the parameters β_i, $i = 1, 2, \ldots, m$ which solves the following problem.

$$\text{Minimize } S = \sum_{j=1}^{n} |y_j - x_{j1}\beta_1 - x_{j2}\beta_2 - \ldots - x_{jm}\beta_m| \tag{1}$$

where $(y_j, x_{j1}, x_{j2}, \ldots, x_{jm})$, $j = 1, 2, \ldots, n$ represents the observed values during n repetitions of an experiment.

The linear programming equivalent of (1) is the following.

$$\text{Minimize } \sum_{j=1}^{n} (P_j + N_j) \tag{2}$$

$$\text{subject to} \quad x_{j1}\beta_1 + x_{j2}\beta_2 + \ldots + x_{jm}\beta_m + P_j - N_j = y_j;\ j = 1, 2, \ldots, n$$

$$P_j, N_j \geq 0$$

where P_j and N_j are the positive and negative deviations of the j-th observation, repectively.

Using matrix notation, we can now write the constraints of LP(2) as follows:

$$X\beta + IP - IN = Y$$

$$P, N \geq 0$$

where $P^T = (P_1, \cdots, P_n)$, $N^T = (N_1, \cdots, N_n)$, $Y^T = (y_1, \cdots, y_n)$, X is an n by m matrix with $n \geq m$, and $\beta^T = (\beta_1, \cdots, \beta_m)$.

3. Multiple Pivoting Techniques of the Special Purpose Algorithms

Our first objective in this section is to present a geometrical interpretation of the L_1 estimation problem. With this explanation, the steps used in the algorithms will be more evident. Geometrically, $S = \sum_{i=1}^{n} |y_i - X_i\beta|$ represents a convex ployhedral surface in Euclidean (m + 1) space with coordinates $(\beta_1, \beta_2, \cdots, \beta_m, S)$ where $X_i = (x_{i1}, x_{i2}, \cdots, x_{im})$. The algorithm considers basic solutions formed by m hyperplanes $X_i\beta = y_i$, $i = 1, 2, \cdots, m$. If β^* is a solution to such a system, then $(\beta^*, S(\beta^*))$ is an extreme point (vertex) on the polyhedral surface. An edge on the polyhedral surface is determined by the intersection of (m - 1) of the hyperplanes $X_i\beta = y_i$. During any iteration, the special purpose algorithms, use m such hyperplanes so that the associated X_i will form a basis for E^m. The intersection of these m hyperplanes determine the current basic solution, and the m edges formed by the possible combinations of these m - 1 hyperplanes are called the basic edges. The algorithms search along the basic edges to determine if a decrease in the sum of the absolute values of the residuals is possible. If a decrease is possible, then an iteration of the algorithm will move along the chosen edge in the prescribed direction by pivoting through the sequence of extreme points on this special path until a decrease no longer occurs. The procedure is repeated until the lowest vertex is reached in the case of a unique solution.

We shall now show how this procedure is implemented in the simplex algorithm. The standard procedure of the simplex algorithm is to choose, first, the vector to enter the basis by picking the most negative marginal (reduced) cost. Then, the vector to leave the basis is determined by the minimum ratio test. The pivoting of the basis vectors then follows. However, the special purpose algorithms do not immediately perform the pivot. These algorithms add twice

the amount of the component of the current tableau associated with the minimum ratio to the reduced cost. If this new reduced cost is positive, then the pivot operation is performed. If the new reduced cost is negative, then the next smallest minimum ratio is considered. Twice the element of the tableau associated with this minimum ratio is then added to the reduced cost. The procedure continues until the reduced cost becomes nonnegative. At that time, the standard pivot procedure is implemented on this last element from the sequence of minimum ratios. At each of these "pivots", a P_i and N_i are interchanged in the basis until the last pivot is performed.

In order to formalize this mathematically, we shall consider LP(2). It follows directly from the LP theory that an optimal basis must contain all m columns of the matrix X and also n - m columns from the matrices I and -I associated with the 2n variables, P_j and N_j. We should note that rank deficiencies are easily treated within the LP framework by the techniques developed for LP theory in Charnes and Cooper (1961). Thus, it is necessary to consider only basis matrices of the form (after row interchanges)

$$B = \begin{pmatrix} X_B & 0 \\ X_R & E \end{pmatrix}$$

where E is a matrix with +1 or -1 on the diagonal and zeroes off the diagonal. The inverse of B is

$$B^{-1} = \begin{pmatrix} X_B^{-1} & 0 \\ -EX_R X_B^{-1} & E \end{pmatrix}$$

The reduced cost for the (n - m) nonbasic $N_i(P_i)$ with corresponding $P_i(N_i)$ in the basis is +2. For the case where both N_i and P_i, m of each, are nonbasic, the reduced cost of P_i is $1 + \sum_{j \varepsilon IR} e_j X_j X^{-1}_{B(k)}$, and for N_i, it is $1 - \sum_{j \varepsilon IR} e_j X_j X^{-1}_{B(k)}$, where $X^{-1}_{B(k)}$ is the k-th column of X_B^{-1}. IR is the index set of X_j's forming X_R, and e_j is the value of the diagonal element of E associated with X_j. The optimality conditions are then $-1 \leq \sum_{j \varepsilon IR} e_j X_j X^{-1}_{B(k)} \leq 1$, k = 1, 2, ..., n.

If the optimality conditions are not satisfied, then it is determined that either N_i or P_i is to enter the basis, and the variable to leave the basis is determined by the minimum ratio test. It is given as follows:

$$\frac{y_j - X_j X_B^{-1} Y_B}{\mu e_j X_j X^{-1}_{B(k)}} \; ; \quad \mu e_j X_j X^{-1}_{B(k)} > 0, \; j \varepsilon IR \tag{5}$$

where Y_B is the vector of the dependent variables corresponding to X_B, and

$$\mu = \begin{cases} -1 \text{ if } P_i \text{ enters the basis} \\ 1 \text{ if } N_i \text{ enters the basis.} \end{cases}$$

In order to illuminate the steps of the multiple pivots, let t be the value of the index IR which yields the minimum ratio. Then the vector to be removed from the basis will be either N_t or P_t. Suppose that it is N_t and that P_i is chosen to enter the basis. If $1 + \sum_{j\varepsilon IR} e_j X_j X^{-1}_{B(k)} + 2e_t X_t X^{-1}_{B(k)} > 0$, then the pivoting procedure is performed. If $1 + \sum_{j\varepsilon IR} e_j X_j X_{B(k)} + 2\, e_t X_t X^{-1}_{B(k)} < 0$, then P_t will replace N_t in the basis, and the reduced cost must be updated. We now return to the minimum ratio test and find the ratio which is next to the smallest in the index set IR. The procedure is continued until the reduced cost becomes nonnegative, and then an iteration is carried out.

4. Criteria for Exploiting Degeneracy

Degeneracy will occur if there are more than m residuals that are zero during an iteration. We shall assume that there are (m+p) residuals that are zero. Geometrically, this will occur when there are (m+p) edges through a vertex on the surface of the convex polytope in E^{m+1}. However, if we consider E^m, then β^* is the point of intersection of (m+p) hyperplanes $X_i\beta = y_i$. Hence the 2(m+p) variables, N_i and P_i, associated with these equations will all be zero, and thus there will be p basic variables which are zero. Theoretically, these p basic variables will not be zero, but they will have a "small" value given by the ε-polynomial when the perturbation technique of Charnes (1952) is applied. The perturbation will eliminate the possibility of cycling of bases, and it will not be necessary to explicitly determine the value of ε.

However, these algorithms do not always proceed along the edge of steepest descent when degeneracy occurs. We use the term, edge of steepest descent, to refer to the segment of an edge which is adjacent to a vertex that will give the greatest rate of decrease in the objective function when moving to an adjacent vertex. We should note that using the multiple pivoting technique on the edge of steepest descent may not give the specialized extreme point path that will provide the greatest decrease in the objective function. There may be another extreme point path that will provide a greater decrease, but this cannot be determined unless iterations of the algorithm are compared.

We shall develop our results on the assumption that it is possible to obtain a decrease in the value of the objective function along one of the basic edges. However, it is important to note that a decrease may not be possible along any of the m basic edges. Therefore, it is possible that several degenerate pivots must be performed in order to obtain a basis which

has an edge along which it is possible to obtain a decrease. This situation may cause the algorithm to cycle if a perturbation technique is not utilized.

We shall now illustrate the method to obtain the edge of steepest descent from the basic edges. Using matrix notation, we can represent the constraints of the linear programming formulation as follows:

$$\begin{pmatrix} X_B & I_B & 0 & -I_B & 0 \\ X_R & 0 & I_R & 0 & -I_R \end{pmatrix} \begin{pmatrix} \beta \\ P_B \\ P_R \\ N_B \\ N_R \end{pmatrix} = \begin{pmatrix} y_B \\ y_R \end{pmatrix}.$$

After multiplying by X_B^{-1}, we obtain the following equations:

$$\beta + X_B^{-1}P_B - X_B^{-1}N_B = X_B^{-1}\, y_B$$

$$-EX_RX_B^{-1}P_B + EP_R + EX_RX_B^{-1}N_B - EN_R = E(y_R - X_RX_B^{-1}y_B).$$

If the solution is not degenerate, then these equations will simplify as below:

$$\beta = X_B^{-1}\, y_B$$

$$E(P_R - N_R) = E(y_R - X_RX_B^{-1}y_B)$$

where either $P_i > 0$ or $N_i > 0$ and $P_i\, N_i = 0$ for all $i \in IR$. If degeneracy occurs during an iteration, then both P_i and $N_i = 0$ for some $i \in IR$. The sum of the absolute values

$$S = \sum_{j=1}^{n} |y_i - X_{j1}\beta_1 \cdots - X_{jm}\beta_m| = eE(y_R - X_R\beta)$$

$$= eE(y_R - X_RX_B^{-1}\, y_B)$$

We shall now assume that degeneracy is present during an iteration. We may note that $P_B = N_B = \vec{0}$ during the iteration because they are nonbasic. For simplicity, we shall assume that $P_{B(i)}$ is to enter the basis. If we slightly increase $P_{B(i)}$, then we can use this to determine the condition for a decrease in the sum of absolute values. We take $P_B^T(\delta) = (0, \ldots, 0, \delta, 0, \ldots, 0)$ where $\delta > 0$ and δ is the ith component of P_B^T. Our equations will now take the form:

$$\beta = X_B^{-1} y_B - X_B^{-1} P_B(\delta) = X_B^{-1} y_B - X_{B(i)}^{-1} \delta$$

$$E(P_R - N_R) = E(y_R - X_R X_B^{-1} y_B) + E X_R X_B^{-1} P_B(\delta)$$

$$= E(y_R - X_R X_B^{-1} y_B) + E X_R X_{B(i)}^{-1} \delta$$

Let ID be the index set of the degenerate variables and $IR^* = IR - ID$. We can now rewrite the above in a partitioned vector notation as:

$$E(P_R - N_R) = \begin{pmatrix} E_{IR^*} (P_R - N_R) \\ \\ E_{ID} (P_R - N_R) \end{pmatrix} = \begin{pmatrix} E_{IR^*} (y_R - X_R X_B^{-1} y_B + X_R X_{B(i)}^{-1} \delta) \\ \\ E_{ID} \; X_R X_{B(i)}^{-1} \delta \end{pmatrix}$$

The new sum of the absolute values is

$$S(\delta) = \delta + \sum_{j \varepsilon IR} e_j (y_i - X_j X_B^{-1} Y_B) + \sum_{j \varepsilon IR^*} e_j X_j X_{B(i)}^{-1} \delta + \sum_{j \; ID} |X_j X_{B(i)}^{-1} \delta|$$

$$= S + \delta(1 + \sum_{j \varepsilon IR^*} e_j X_j X_{B(i)}^{-1}) + \sum_{j \varepsilon ID} | X_j X_{B(i)}^{-1} \delta|$$

Thus to have the condition that $S(\delta) < S$, we must have

$$1 + \sum_{j \varepsilon IR^*} X_j X_{B(i)}^{-1} + \sum_{j \varepsilon ID} |X_j X_{B(i)}^{-1}| < 0 \qquad \text{(I)}$$

Similarly, if $N_{B(i)}$ were to be chosen to be increased, the criterion would be

$$1 - \sum_{j\varepsilon IR^*} X_j X^{-1}_{B(i)} + \sum_{j\varepsilon ID} |X_j X^{-1}_{B(i)}| < 0. \qquad \text{(II)}$$

By using this new criterion for basis entry, the minimum ratio rule is thus modified to the smallest positive ratio to determine the vector to leave the basis. The procedure for performing the multiple pivots will remain the same. There will no longer be any degenerate pivots, and the objective function will strictly decrease at each iteration. However, examples can be constructed where use of this criterion alone will indicate that a basic edge with a negative rate of change does not exist, yet the corresponding extreme point is not optimal. Thus the standard LP criterion should be used to verify optimality.

5. Examples

In the first example, we shall illustrate the difficulty caused by degeneracy when using the standard simplex criteria for choosing the vector to enter the basis. We shall then present the sequence of tableaus when the new criteria for selection is applied. The multiple pivots will be denoted by asteriks. The simplex format will consider LP(2) written in equivalent maximization form, and it will follow a combination of the notation of Davies, and Barrodale and Roberts. For brevity, the revised simplex tableaus will be used. The problem is to find the LAV estimators for the model $y = \beta_1 + \beta_2 X$ from the following data:

x	-5	-3	-3	-2	0	1	2	2	3	4	4
y	0	2	-3	-1	0	1	3	-1	2	0	4

The initial basis matrix X_B will consist of the sixth and eleventh observations, and degeneracy will occur in this initial tableau. The rule which we used to determine the assignment of N_i or P_i as the degenerate basic variable in this initial tableau was to use the sign of the y_i from the data, i. e., since $y_3 = -3$, we assigned N_3 as the degenerate basic variable, but for $y_5 = 0$, we arbitrarily assigned P_5. Some algorithms use the assignment of a P_i for each degenerate value that occurs. This rule can also lead to computational problems. The sequences of pivots are quite easy to follow when a graph of the parameter (β_1, β_2) space is used. Finally, the optimal tableaus also indicate that alternate optimal solutions are available.

TABLEAU I				TABLEAU II			
BASIS	RESIDUAL	N_6	P_{11}	BASIS	RESIDUAL	P_4	P_{11}
P_1	5	3	2	P_1	21/6	9/6	3/6
P_2	5	7/3	4/3	P_2	23/6	7/6	1/6
N_3	0	-7/3	-4/3	N_3	7/6	-7/6	-1/6
P_4	1	2^{**}	1	N_6	3/6	-3/6	$3/6^{*}$
P_5	0	$4/3^{*}$	1/3	N_5	4/6	-4/6	2/6
β_1	0	-4/3	-1/3	β_1	4/6	-4/6	2/6
P_7	1	2/3	1/3	P_7	4/6	2/6	-4/6
N_8	3	-2/3	1/3	N_8	20/6	-2/6	4/6
N_9	1	-1/3	2/3	N_9	7/6	-1/6	$5/6^{**}$
N_{10}	4	0	1	N_{10}	24/6	0	1
β_2	1	1/3	1/3	β_2	5/6	1/6	1/6
	20	-5	-4		18 5/6	5/6	-13/6

TABLEAU III

BASIS	RESIDUAL	N_4	P_9
P_1	14/5	8/5	3/5
P_2	18/5	6/5	1/5
N_3	7/5	-6/5	-1/5
P_6	1/5	2/5*	-3/5
N_5	1/5	-3/5	2/5
β_1	1/5	3/5	2/5
P_7	8/5	1/5	-4/5
N_8	12/5	-1/5	4/5
P_{11}	7/5	-1/5	-6/5
N_{10}	13/5	1/5	6/5
β_2	3/5	-1/5	1/5
	16 1/5	-2/5	3/5

TABLEAU IV - OPTIMAL

BASIS	RESIDUAL	P_6	P_9
P_1	2	-4	3
P_2	3	-3	2
N_3	2	3	-2
N_4	1/2	5/2	-3/2
N_5	1/2	3/2	-1/2
β_1	1/2	3/2	-1/2
P_7	3/2	-1/2	-1/2
N_8	5/2	1/2	1/2
P_{11}	3/2	1/2	-3/2
N_{10}	5/2	-1/2	3/2
β_2	1/2	-1/2	1/2
	16	1	0

If we return to Tableau I and apply our new criteria (I) and (II) to find the "nondegenerate" rate of change (of decrease) in the objective function, we shall now obtain an opportunity cost of -7/3 for N_6 and -10/3 for P_{11}. Upon observing Tableau II, we find that the objective function value is $20 - 7/3 \cdot 1/2 = 18\ 5/6$, N_3 has increased from 0 to $|-7/3 \cdot 1/2| = 7/6$, and N_5 increased from 0 to $|-4/3 \cdot 1/2| = 4/6$. These would have the same values (and the rest of the tableau) if only the nondegenerate pivot had been performed. However, since P_{11} has the most negative reduced cost, then its associated edge gives the greatest rate of decrease among the basic vectors. The (nondegenerate) multiple pivots are given in Tableau IA. The following tableau will be optimal and identical to Tableau IV.

TABLEAU IA

BASIS	RESIDUAL	N_6	P_{11}
P_1	5	3	2
P_2	5	7/3	4/3
N_3	0	-7/3	-4/3
P_4	1	2	1^*
P_5	0	4/3	1/3
β_1	0	-4/3	-1/3
P_7	1	2/3	-1/3
N_8	3	-2/3	1/3
N_9	1	-1/3	$2/3^{**}$
N_{10}	4	0	1
β_2	1	1/3	1/3
	20	-7/3	-10/3

Modifying the original data, we shall illustrate the possibility of not obtaining a decrease along the basic edges. We require an additional observation, the twelfth, of $x = 1/2$, $y = 0$ for our data. The initial matrix X_B will consist of the tenth and twelfth observations, and degeneracy will occur in Tableau V. This is not the final tableau as the standard reduced costs of the simplex algorithm are not positive. Thus a degenerate pivot is performed, and using the perturbation technique of Charnes (1952), we find that P_5 and N_{12} will be exchanged in X_B. For this particular example, the replacement of the twelfth observation by either the first or the fifth observation in X_B will give a basic edge along which the objective function will decrease. It should also be noted that only a single degenerate pivot will be executed since the addition of twice the component of the tableau of either of the tied minimum ratios to the reduced cost will make it positive.

At this point, it will be very beneficial to examine our criteria which gives the actual rate of change of the objective function along the basic edges. The rates of changes for P_{10}, P_{12}, and N_{12} are, respectively 37/7, 1/7, 29/7, and 37/7. Since these values are all positive, then proceeding in any direction along these basic edges would increase the sum of the absolute values. To demonstrate this, suppose we select N_{10} to be increased. The first minimum ratio is associated with N_4, and it is 7/5. X_B will consist of the fourth and twelfth observations when only this pivot is accomplished. The point will be $\beta_1 = -1/5$, $\beta_2 = 2/5$, and the sum of the absolute values will be 17 1/5. We can determine the change in the objective function by using the information provided by the algorithm. Thus, we have 17 1/5 = 17 + (1/7) (7/5).

We shall conclude our discussion of this example by stating that optimality will be reached in another two tableaus after the interchange of P_5 and N_{12} in X_B. The optimal solution is $\beta_1 = 1/5$, $\beta_2 = 3/5$, and the sum of the absolute values is 16 7/10. X_B will consist of the fourth and nineth observations.

Modifying this example, we shall illustrate the possibility of not obtaining a decrease along the initial basic edges. We require an additional observation, the twelfth, of X = 1/2, y = 0 for our data. The initial matrix X_B will consist of the tenth and twelfth observations, and degeneracy also occurs in this tableau (V). It should be clear that the tableau is not optimal and that the addition of twice the component of the tableau of either of the tied minimum ratioes to the reduced cost will make it positive. For this particular example, the replacement of the twelfth observation by either the first or fifth observation in X_B will give a basic edge along which the objective function will decrease. The optimal solution is $\beta_1 = 1/5$, $\beta_2 = 3/5$, and X_B consists of the fourth and nineth observations.

TABLEAU V

BASIS	RESIDUAL	P_{10}	N_{12}
P_1	0	11/7	18/7
P_2	2	7/7	14/7
N_3	3	-7/7	-14/7
N_4	1	-5/7	-12/7
P_5	0	1/7	8/7
P_6	1	-1/7	6/7
P_7	3	-3/7	4/7
N_8	1	3/7	-4/7
P_9	2	-5/7	2/7
β_1	0	-1/7	-8/7
P_{11}	4	-7/7	0
β_2	0	2/7	2/7
	17	13/7	-15/7

6. Degeneracy and Optimality

Although we have taken advantage of the special structure of the linear programming formulation and its associated geometrical considerations to find a condition that should improve convergence when degeneracy occurs during an iteration, we must emphasize the fact that our criteria (I) and (II), in general, cannot be used as sufficient conditions for optimality. The last example in the preceding section clearly demonstrates that the positivity of (I) and (II) does not imply optimality. However, if the objective function has assumed its minimum value, then the criteria (I) and (II) will be nonnegative, i.e., the conditions are necessary. Supplemental conditions will now be developed which are sufficient.

Conditions (I) and (II) may be viewed as being derived from a dynamic perturbation of the problem. This perturbation forces for $j \in ID$, either N_j or P_j into the basis depending on whether the sign of $X_j X^{-1}_{B(i)}$ is positive or negative. Hence, movement along a basic edge will always be "away from" other nonbasic hyperplanes passing through the point β^*. To some extent this dynamic perturbation is theoretically justifiable because the perturbation as defined by Charnes (1952) depends only on the constraint index and this ordering is arbitrary. After a nondegenerate pivot the perturbation may be redefined without the danger of cycling. The difficulty with (I) and (II) is that a different perturbation may be enforced at a <u>single</u> iteration. This means that for each j, with X_j a row of X_B, we may be looking at the reduced costs for different problems. If there are p degeneracies at an iteration then without varying X_B there are 2^p different bases of the form (1) for this extreme point. A general representation of the reduced costs for these 2^m bases is:

$$1 + \sum_{j\epsilon IR^*} e_j X_j X^{-1}_{B(i)} - \sum_{j\epsilon ID} d_j X_j X_{B(i)} \qquad \text{(III)}$$

$$1 - \sum_{j\epsilon IR^*} e_j X_j X^{-1}_{B(i)} - \sum_{j\epsilon ID} d_j X_j X^{-1}_{B(i)} \qquad \text{(IV)}$$

for P_k and N_k, respectively, where X_k is the i-th row of X_B and d_j = +1 or -1.

The current solution is optimal when a set of d_j $j\epsilon ID$ can be found such that (III) and (IV) are nonnegative for i = 1, 2, ...m. Maximizing the minimum value of (III) and (IV) over i = 1, 2, ...m. is an integer programming problem. It would not appear to be computationally beneficial to solve this problem; however, certain d_j are readily available from the computation of (I) and (II). These can be conveniently used to verify optimality in certain instances.

7. Remarks on Computation Results

We modified a multiple pivoting code developed by Armstrong and Frome (1976) to include our criteria for finding the basic edge of steepest descent. The algorithm was coded in Fortran, and the criteria were implemented in a straightforward manner. Our objectives were to determine if the number of pivots and the execution time would be decreased. We chose to compare the results of our coded algorithm with the original code of Armstrong and Frome. We generated our data from discrete distributions since, theoretically, degeneracy will not occur in the case of continuous distributions. In most of the problems in which there were only a few (4 or less) degenerate variables at an iteration, our code required slightly less pivots than the original code, but our execution time exceeded theirs. In most problems

with several (or many) degenerate variables at an iteration, we found our algorithm showed a significant improvement in the number of iterations so compared to the other code. However, in terms of execution time, our algorithm showed only a slight improvement in speed of convergence in some problems, and in other problems, it was slower in reaching optimality.

8. Observations

The purpose of this paper was to illustrate that degeneracy in LAV estimation can increase the number of iterations and also solution time in algorithms which have been shown to be very efficient. In problems where a large number of observations are taken to estimate a few parameters, it is very likely that degeneracy will occur. We have been able to develop an improvement in the multiple pivoting algorithms when degeneracy occurs which enables the pivots to proceed along the basic edge which gives the greatest rate of decrease in the objective function. The results of our computational study verify our hypothesis on decreasing the number of iterations. The current direction of study involves a more sophisticated implementation of the criteria into the code. One possible direction in which an extension of the procedure would be desirable is to have an efficient method of determining the rate of decrease for all edges at a point where degeneracy is present. However, to pivot through all possible bases by the simplex algorithm to find the rates would require c(m+k,m) pivots, and this would not be advisable.

REFERENCES

Armstrong, R. D. and Frome, E. L. (1976). "A Comparison of Two Algorithms for Algorithms for Absolute Deviation Curve Fitting," Journal of the American Statistical Association, 71, 328-330

Barrodale, I. and Roberts, F. D. K. (1973). "An improved algorithm for discrete L_1 linear approximation," SIAM J. Numerical Analysis, 10, 839-848.

Charnes, A. (1952). "Optimality and degeneracy in linear programming," Econometrica, 20, 160-170.

Charnes, A. and Cooper, W. W., (1961). Management Models and Industrial Applications of Linear Programming, Vol. I and II, John Wiley and Sons, New York.

Charnes, A., Cooper, W. W., and Ferguson, R. (1955). "Optimal estimation of executive compensation by linear programming," Management Science, 2, 138-151.

Darboux, G., Ed. (1888-90). Oeuvres de Fourier, Gauthier-Villars, Paris, 2, 324-325.

Davies, M. (1967). "Linear approximation using the criterion of least total deviations," J. R. Statist. Soc B, 29, 101-109.

Edgeworth, F. Y. (1887). "On discordant observations," Philosophical Magazine, 24, 264-375.

Edgeworth, F. Y. (1888). "On a new method of reducing observations relating to several quantities," Philosophical Magazine, 5th Series, 25, 185-191.

Edgeworth, F. Y. (1923). "On the use of medians for reducing observations relating to several quantities," Philosophical Magazine, 6th Series, 1074-1088.

Poussin, M. Ch. J. de la Vallée, (1911). "Sur la méthode de l'approximation minimum," Ann. Soc. Sci. de Bruxelles, 35, 1-16.

Rhodes, E. C. (19930). "Reducing observations by method of minimum deviations," Philosophical Magazine, 7th Series, 954-992.

Singleton, R. R. (1940). "A method for minimizing the sum of absolute values of deviations," Annals Math. Statist., 11, 301-310.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Center for Cybernetic Studies
The University of Texas

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
6 Degeneracy in Special Purpose Primal Algorithms Used in Cbtaining Least Absolute Value Estimators.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
9 Research rept.,

5. AUTHOR(S) (First name, middle initial, last name)
10 R. D. Armstrong, J. P. Godfrey
A. Charnes

6. REPORT DATE
11 Nov[redacted]77 12 26p.

7a. TOTAL NO. OF PAGES
23

7b. NO. OF REFS
13

8a. CONTRACT OR GRANT NO.
N00014-75-C-0616; 0569
NSF Grant MCS 77-01000

b. PROJECT NO.
NR 047-021

d. 15 N00014-75-C-0616,
N00014-75-C-0569

9a. ORIGINATOR'S REPORT NUMBER(S)
Center for Cybernetic Studies
Research Report CCS-316
14

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT
This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Cffice of Naval Research (Code 434)
Washington, D. C.

13. ABSTRACT

Efficient algorithms have been developed recently which utilize the specialized structure of the linear programming formulation for the problem of least absolute value estimation. These algorithms generally proceed in the direction of steepest descent along an edge of a convex polyhedral surface. However, we will show that the extreme point path of steepest descent may not be taken when degeneracy occurs. We will also present a criterion that determines the basic edge for steepest descent.

DD FORM 1 NOV 65 1473 (PAGE 1)
S/N 0101-807-6811

406197

Unclassified
Security Classification

A-31408

CL

Unclassified

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
L_1 Norm						
Least Absolute Value						
Linear Programming						